

Local stability in a transient Markov chain

Ivo Adan^{*} Sergey Foss[†] Seva Shneer[‡] Gideon Weiss[§]

November 20, 2015

Abstract

We prove two lemmas with conditions that a system, which is described by a transient Markov chain, will display local stability. Examples of such systems include partly overloaded Jackson networks, partly overloaded polling systems, and overloaded multi-server queues with skill based service, under first come first served policy.

Keywords: Markov chains, Local stability, Jackson networks, polling systems, skill based service

1 Introduction

Many complex stochastic systems can be described by an irreducible Markov chains on a countable state space. It is often the case that the state of this Markov chain is composed of several components, where each component describes the “local” state of part of the system. While the dynamics of the system, given by the transition mechanism of the Markov chain, is influenced by the state of the entire system, it is often the case that the dependence of the local transitions is only weakly coupled with the rest of the system.

Essential to the study of Markov chains is the question of stability: Is the chain ergodic, in which case it has a stationary distribution from which its long time average behavior can be obtained, or is it transient, in which case it may be studied through fluid approximations. These two modes of behavior are totally different.

In complex systems one may however be faced by an intermediate sort of behavior. While the system as a whole is transient, and so there is no stationary distribution for the entire system, some components of the system, when regarded locally, display stable behavior and seem to approach a stationary distribution when regarded on their own, at least for most of the time.

We formulate this type of situation, and prove that under the proper conditions one can indeed talk about local stability of such transient systems. This is done in Lemmas 1 and 2, in Section 3. Before that, in Section 2, we present some examples.

^{*}Department of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, the Netherlands; email iadan@tue.nl Research supported in part by the Netherlands Organization for Scientific Research (NWO).

[†]Heriot Watt University, Edinburgh and Novosibirsk State University; email s.foss@hw.ac.uk Research supported in part by EPSRC grant No. EP/I017054/1

[‡]Heriot-Watt University, Edinburgh; email v.shneer@hw.ac.uk Research supported in part by EPSRC grant No. EP/L026767/1

[§]Department of Statistics, The University of Haifa, Mount Carmel 31905, Israel; email gweiss@stat.haifa.ac.il Research supported in part by Israel Science Foundation Grant 711/09 and 286/13, and by Netherlands STAR grant.

2 Examples of local stability

2.1 Jackson networks

This example is analyzed by Goddman and Massey [11]. In a Jackson network, single server nodes $i = 1, \dots, I$ have exogenous arrival rate ν_i , service rate μ_i , and routing probabilities $P_{i,j}$ for a customer that complete service at node i to go next to node j , with the matrix P sub-stochastic with spectral radius < 1 . The traffic equations for a stable Jackson network are:

$$\lambda = \nu + P' \lambda$$

with λ_i the stationary rate of inflow and outflow of customers from node i . Necessary and sufficient for ergodicity is $\lambda_i < \mu_i$, i.e. $\rho_i = \lambda_i / \mu_i < 1$, for $i = 1, \dots, I$. Assuming Poisson arrivals and exponentially distributed services the stationary distribution of the queue lengths $Q(t)$ is then of product form:

$$P(Q_i(t) = n_i, i = 1, \dots, I) = B \prod_{i=1}^I \rho_i^{n_i}$$

where B is the normalising constant. If the ergodicity condition does not hold for all the nodes, then the system is transient. The traffic equations are now modified to

$$\lambda = \nu + P'(\lambda \wedge \mu),$$

which has a unique solution, reached by solving a linear complementarity problem. It divides the nodes into two sets: $I_0 = \{i : \lambda_i < \mu_i\}$ and $I_1 = \{i : \lambda_i \geq \mu_i\}$, and the following holds:

At the nodes for which $\lambda_i > \mu_i$, the queue lengths diverges to infinity, with

$$\frac{Q_i(t)}{t} \rightarrow \lambda_i - \mu_i, \text{ as } t \rightarrow \infty.$$

At the nodes for which $\lambda_i < \mu_i$, the queue lengths converge in law to the product form distribution

$$\mathbf{P}(Q_i(t) = n_i, i \in I_0) \rightarrow B \prod_{i=1}^I \rho_i^{n_i}, \text{ as } t \rightarrow \infty \quad (1)$$

and this convergence holds irrespective of the initial state of the network.

At the nodes for which $\lambda_i = \mu_i$ one has

$$\mathbf{P}(Q_i(t) = n_i) \rightarrow 0, \text{ as } t \rightarrow \infty,$$

but

$$\mathbf{P}(Q_i(T_k) = 0 \text{ for infinitely many } T_k) = 1,$$

and the time distances $T_{k+1} - T_k$ have infinite mean.

Here nodes in I_1 diverge (are unstable), while all the nodes in $i \in I_0$ act like a Jackson network with augmented input, where for each i , in addition to exogenous input rate ν_i , there is input from each node $j \in I_1$, at rate $\mu_j P_{j,i}$.

The difficulty in verifying (1) is that there is no steady state distribution for the whole system since it is transient, and there are no equilibrium equations for the subsystem of node I_0 , since

they do not form a Markov chain. Goodman and Massey prove (1) by considering two ergodic Markov chains for I_0 which provide stochastic upper and lower bound for $Q_i(t)$, $i \in I_0$, and one of which has stationary distribution (1), while the other has a parameter ϵ , and stationary distribution which converges to (1) as $\epsilon \rightarrow 0$.

What we observe here is that clearly $Q_i(t)$, $i \in I_1$ must diverge and therefore will be > 0 from some point in time onwards (at least for those i for which $\lambda_i > \mu_i$). Furthermore, when $Q_i(t) > 0$, for $i \in I_1$ then node i provides a Poisson input stream of jobs entering the other nodes at the constant rate μ_i , so given $Q_i(t) > 0$, $i \in I_1$, the system of nodes $i \in I_0$ does indeed behave like a Jackson network with the augmented input.

2.2 Multi-server queues with skill based service under FCFS policy

This example is analyzed by Adan and Weiss [2]. In a skill based service queue there are customers of types $\mathcal{C} = \{c_1, \dots, c_I\}$ and servers $\mathcal{S} = \{s_1, \dots, s_J\}$, and a bipartite compatibility graph between \mathcal{S} and \mathcal{C} , with an arc (s_j, c_i) if server s_j can serve customers of type c_i . Assume customer arrivals are Poisson at rates λ_{c_i} , and service is exponential, with rates μ_{s_j} , let $\lambda = \sum_{\mathcal{C}} \lambda_{c_i}$, $\mu = \sum_{\mathcal{S}} \mu_{s_j}$, $\alpha_{c_i} = \lambda_{c_i}/\lambda$, $\beta_{s_j} = \mu_{s_j}/\mu$. Denote $\mathcal{S}(c_i)$ the servers of c_i , $\mathcal{C}(s_j)$ the customers of s_j , and for $C \subseteq \mathcal{C}$, $S \subseteq \mathcal{S}$, let $\mathcal{S}(C) = \bigcup_{c_i \in C} \mathcal{S}(c_i)$, $\mathcal{C}(S) = \bigcup_{s_j \in S} \mathcal{C}(s_j)$, and let also $\mathcal{U}(S) = \overline{\mathcal{C}(\overline{S})}$ be customer types which can only be served by servers in S . Let $\alpha_C = \sum_{c_i \in C} \alpha_{c_i}$ and $\beta_S = \sum_{s_j \in S} \beta_{s_j}$, with analogous notation for λ_C , μ_S .

Service discipline is first come first served (FCFS) assign longest idle server (ALIS), i.e. server s_j , when free, will take the longest waiting compatible customer, and arriving customer of type c_i is assigned to longest idle compatible server.

The following Markov chain $X(t)$ describes this system: imagine the customers ordered by order of arrivals, with the busy servers positioned at the location of the customers which they are serving, and the idle servers located after the last customer, ordered by increasing idle time. The state of the Markov chain is given by the random permutation of the servers, and by the lengths of the queues between the servers, where we write $\mathfrak{s} = (S_1, n_1, \dots, S_i, n_i, S_{i+1}, \dots, S_J)$ for the state where servers S_1, \dots, S_i are busy, with S_1 serving the earliest customer, and n_j customers waiting between S_j and S_{j+1} , and S_i is the last busy server, with n_i customers behind it, and servers S_{i+1}, \dots, S_J are idle ordered by length of idle time, with S_J longest idle.

This Markov chain is ergodic if and only if for every non-empty subset of customer types C , and of servers S , the three equivalent sets of conditions hold:

$$\lambda_C < \mu_{\mathcal{S}(C)}, \quad \mu_S < \lambda_{\mathcal{C}(S)}, \quad \mu_S > \lambda_{\mathcal{U}(S)}. \quad (2)$$

In that case the stationary distribution of the Markov chain is given by:

$$\pi(\mathfrak{s}) = B \prod_{j=1}^i \frac{\lambda_{\mathcal{U}(\{M_1, \dots, M_j\})}^{n_j}}{\mu_{\{M_1, \dots, M_j\}}^{n_j+1}} \prod_{j=i+1}^J \lambda_{\mathcal{C}(\{M_j, \dots, M_J\})}^{-1}. \quad (3)$$

with normalizing constant B .

In particular, if the condition

$$\alpha_C < \beta_{\mathcal{S}(C)}, \quad \beta_S < \alpha_{\mathcal{C}(S)}, \quad \beta_S > \alpha_{\mathcal{U}(S)} \quad C \neq \mathcal{C}, \emptyset, \quad S \neq \mathcal{S}, \emptyset \quad (4)$$

then the system is ergodic for all $\lambda < \mu$. The condition (4) is referred to as *Complete Resource Pooling*.

If condition (2) fails, then $X(t)$ is transient, and queues of some types of customers will grow to infinity. However, local stability, in the sense that servers stay close together, may still hold.

Assume that $\lambda > \mu$ but complete resource pooling condition (4) holds. For state $X(t) = \mathbf{s} = (M_1, n_1, \dots, M_i, n_i, M_{i+1}, \dots, M_J)$ let $k(t) = i$, $Q_j(t) = n_j$, $j = 1, \dots, k(t)$. Then, as $t \rightarrow \infty$:

$$k(t) \rightarrow J \text{ and } \frac{Q_J(t)}{t} \rightarrow \lambda - \mu \text{ a.s.}, \quad (5)$$

$$\mathbf{P}(M_1, n_1, \dots, M_{J-1}, n_{J-1}, M_J, n_J > 0) \rightarrow B_0 \prod_{j=1}^J \frac{\alpha_{\mathcal{U}(\{M_1, \dots, M_j\})}^{n_j}}{\beta_{\{M_1, \dots, M_j\}}^{n_j+1}}. \quad (6)$$

In words, while the queue behind the last server grows without limit, all the servers are busy all the time, and their permutation and the number of customers waiting between them converges to a limiting distribution. This limiting distribution is in fact the stationary distribution of an ergodic model, of FCFS infinite bipartite matching model discussed in [1, 3].

If the queue does not have complete resource pooling, i.e. (4) fails, then there is a unique decomposition of the customer types and of the servers into subsets which are constructed recursively, for $i = 1, \dots, L$ as follows:

$$\mathcal{C}^{(i)} = \arg \min_{C \subseteq \mathcal{C} \setminus \bigcup_{k < i} \mathcal{C}^{(k)}} \frac{\beta_{\mathcal{S}(C)}}{\alpha_C}, \quad \mathcal{S}^{(i)} = \mathcal{S}(\mathcal{C}^{(i)}).$$

and each of these subsystems on its own has complete resource pooling. With these subsystem are associated values $0 = \lambda^{(0)} < \lambda^{(1)} < \dots < \lambda^{(L)} < \lambda^{(L+1)} = \infty$, so that for $\lambda^{(i)} < \lambda < \lambda^{(i+1)}$, the subsystems $\mathcal{C}^{(l)}, \mathcal{S}^{(l)}$, $l = 1, \dots, i$ converge in law to a stationary limiting distribution analogous to (6), while the remaining servers and customer types will converge to a limiting distribution analogous to (3), and for $l = 1, \dots, i$ the queue between the last server of $\mathcal{S}^{(l)}$ and the first server of $\mathcal{S}^{(l+1)}$ will grow to infinity, as $t \rightarrow \infty$.

2.3 Mesh network governed by a CSMA/CA protocol

In this example we will look at the performance of a mesh network on a line where transmissions of nodes are governed by the CSMA/CA protocol. Let us consider one of the (slightly) different models of such a network described in [12] (see also [4], [5], [7], [13]). Namely, consider a random-access network consisting of n nodes on a line, numbered $1, 2, \dots, n$. Each node acts as a transmitter and a receiver and is assumed to have an infinite buffer for storing messages. Every node transmits messages to the next node on its right. A message enters the system through the left-most node, needs to be relayed by all nodes and leaves the system once it has been transmitted by the right-most node.

We say that nodes within distance k are neighbours, and are prevented from transmitting simultaneously.

Time is slotted, and the transmission time of any message by any node is assumed to be equal to the duration of the slot. It is then assumed that each time slot may be partitioned into a contention period and a data period. At the beginning of a contention period all nodes (with non-empty buffers) draw a uniformly distributed back-off time between 0 and the length of the contention period. A node activates when its back-off timer runs out, but only if no nodes within distance k are already active. Nodes then transmit for the entire duration of the data period. The duration of the contention period has to be sufficiently large to allow the carrier-sensing mechanism to function correctly. However, it can always be assumed to be much smaller than the length of the data period by scaling up the transmission durations. Without loss of generality, we will assume the length of the contention period to be zero. Therefore we can view the competition between nodes for transmitting as follows: among all the nodes with a message to transmit, at the beginning of a time slot an order of priorities is chosen at random uniformly

among all possible orders. Then a node will transmit a message in this time slot if and only if its priority is higher than that of any of its neighbours.

A first question of interest when studying such systems is the end-to-end throughput (average number of messages per time slot leaving the system) if the first (left-most) node has an infinite supply of messages. To answer that, one needs to understand the long-term behaviour of the queues of all nodes. In [12] the authors study the simplest case when $n = 2k + 1$ and show that in this case the queues of the first $k + 1$ nodes tend to infinity, while the queues of other nodes remain bounded. Heuristically this is clear as for any of the first $k + 1$ nodes, it has exactly one more competitor than its immediate neighbour to the left and will therefore transmit messages less frequently than receive. The situation is opposite for the nodes from $k + 2$ to n : each node has exactly one competitor less than its immediate neighbour to the left (we refer the reader to a more rigorous, although still incomplete, treatment in [12]).

There are a number of open questions. The first one is of course how to make the statements of [12] complete and rigorous. This has been done recently in [13] for the simplest network with $n = 3$ and $k = 1$. It would also be interesting to consider cases of general n and k . A further important relaxation one can make is to assume that there is an arrival stream of intensity, say, λ into the left-most node - it would be interesting to look at the increasing λ and investigate which queues become saturated and how.

2.4 Polling Systems

Foss, Chernova and Kovalevskii [10] consider a polling system with one or several servers, and stations $k = 1, \dots, K$. Input consists of stationary ergodic streams of customers. Servers follow i.i.d. cyclic routes through all the stations, which are independent of the arrivals and of the queues $Q_k(t)$. Service policy has a number $f_k^j(x, D_k^j)$ of customers served on the j th visit of the server to station k , if there are x customers at the station, where D_k^j are i.i.d., and service is monotone in the sense that $f_k(x, D) \leq f_k(x + 1, D) \leq f_k(x, D) + 1$.

In these systems it is possible that some of the queues, say at stations $k \in K_1$ are unstable while the remaining queues, at stations $k \in K_0$ display local stability.

3 Lemmas on local stability for a transient Markov chain

Lemma 1. *Let $X(n) = (X_1(n), X_2(n))$ be a Markov chain on countable state space with $X_1(n) = (X_{11}, \dots, X_{1,k}) \in \mathbb{Z}_+^k$ and $X_2 = (X_{21}, \dots, X_{2m}) \in \mathbb{Z}_+^m$. Assume the following:*

1. $\lim_{n \rightarrow \infty} X_{2i}(n) = \infty$ almost surely, for all $i = 1, \dots, m$ and for any initial condition $(X_1(0), X_2(0))$.
2. $P(X_1(n+1) = j | X_1(n) = i, X_2(n) = l) = P_{i,j}$, for all values of $l = (l_1, \dots, l_m)$ with all strictly positive coordinates, where $P_{i,j}$ are transition probabilities of an ergodic Markov chain with the unique stationary distribution $\pi = \{\pi_j\}$.

Then for all initial i_0, j_0 :

$$\sup_j \left| \mathbf{P}(X_1(n) = j | X_1(0) = i_0, X_2(0) = j_0) - \pi_j \right| \rightarrow 0, \text{ as } n \rightarrow \infty$$

i.e. $X_1(n)$ converges in distribution to π in the total variation norm.

Proof. To simplify the notation, we provide a proof for $k = m = 1$. Denote the transition probabilities of $X(t)$ as follows:

$$\mathbf{P}(X_1(n+1) = k, X_2(n+1) = l \mid X_1(n) = i, X_2(n) = j) = P_{(i,j),(k,l)}$$

By property (2) we can write, whenever $j > 0$:

$$\begin{aligned} & \mathbf{P}(X_2(n+1) = l \mid X_1(n+1) = k, X_1(n) = i, X_2(n) = j) \\ &= \frac{\mathbf{P}(X_1(n+1) = k, X_2(n+1) = l \mid X_1(n) = i, X_2(n) = j)}{\mathbf{P}(X_1(n+1) = k \mid X_1(n) = i, X_2(n) = j)} \\ &= \frac{P_{(i,j),(k,l)}}{P_{i,k}} \end{aligned}$$

Fix initial values $X_1(0) = i_0, X_2(0) = j_0$, and choose arbitrary $\epsilon > 0$. Choose large enough n_0 (to be specified). Starting at time n_0 with the values $X_1(n_0), X_2(n_0)$, we now construct a new Markov chain, $(Y(n), \hat{X}_1(n), \hat{X}_2(n))$ for $n = n_0, n_0 + 1, \dots$. We initialize them at time n_0 as:

$$\begin{aligned} Y(n_0) &= \hat{X}_1(n_0) = X_1(n_0), \\ \hat{X}_2(n_0) &= X_2(n_0), \end{aligned}$$

Starting from these values at n_0 the following transitions are made, from n to $n+1$ for $n \geq n_0$:

$$\begin{aligned} \mathbf{P}(Y(n+1) = k \mid Y(n) = i) &= P_{i,k} \\ \hat{X}_1(n+1) &= \begin{cases} Y(n+1) & \text{if } \hat{X}_2(n) > 0 \\ * & \text{if } \hat{X}_2(n) = 0 \end{cases} \end{aligned}$$

and the value of $\hat{X}_2(n+1)$ is generated as follows:

$$\begin{aligned} \mathbf{P}(\hat{X}_2(n+1) = 0 \mid Y(n), \hat{X}_1(n), \hat{X}_2(n) = 0) &= 1 \\ \mathbf{P}(\hat{X}_2(n+1) = l \mid \hat{X}_1(n+1) = k, \hat{X}_1(n) = i, \hat{X}_2(n) = j, j > 0) &= \frac{P_{(i,j),(k,l)}}{P_{i,k}} \end{aligned}$$

Observe that $Y(n)$ on its own is a Markov chain, with transition probabilities $P_{i,k}$. Also, $\hat{X}_1(n), \hat{X}_2(n)$ on its own is distributed for $n \geq n_0$ exactly like $X_1(n), X_2(n)$, for as long as $\hat{X}_2(n-1) > 0$. Once $\hat{X}_2(n) = 0$, it will stay as 0 for all times $m \geq n$, and $\hat{X}_1(m) = *$ for all $m > n$. Finally, note that $\hat{X}_1(n) = Y(n)$ for as long as $\hat{X}_2(n-1) > 0$.

The following holds:

$$\mathbf{P}(X_1(n) = j \text{ and } X_2(m) > 0 \text{ for all } n_0 \leq m \leq n-1) = \mathbf{P}(\hat{X}_1(n) = j) = P(\hat{X}_1(n) = Y(n) = j)$$

We now look at the total variation distance between the distribution of $X_1(n)$ (having started

from the fixed i_0 at time 0), and the distribution π .

$$\begin{aligned}
|\mathbf{P}(X_1(n) = j) - \pi_j| &= |\mathbf{P}(X_1(n) = j \text{ and } X_2(m) > 0 \text{ for all } n_0 \leq m \leq n-1) \\
&\quad + \mathbf{P}(X_1(n) = j \text{ and } X_2(m) = 0 \text{ for some } n_0 \leq m \leq n-1) - \pi_j| \\
&= |\mathbf{P}(\widehat{X}_1(n) = Y(n) = j) + \mathbf{P}(X_1(n) = j \text{ and } X_2(m) = 0 \text{ for some } n_0 \leq m \leq n-1) - \pi_j| \\
&= |\mathbf{P}(Y(n) = j) - \pi_j - \mathbf{P}(Y(n) = j, \widehat{X}_1(n) \neq Y(n)) \\
&\quad + \mathbf{P}(X_1(n) = j \text{ and } X_2(m) = 0 \text{ for some } n_0 \leq m \leq n-1)| \\
&\leq |\mathbf{P}(Y(n) = j) - \pi_j| + \mathbf{P}(Y(n) = j, \widehat{X}_1(n) \neq Y(n)) \\
&\quad + \mathbf{P}(X_1(n) = j \text{ and } X_2(m) = 0 \text{ for some } n_0 \leq m \leq n-1) \\
&\leq |\mathbf{P}(Y(n) = j) - \pi_j| + 2\mathbf{P}(X_2(m) = 0 \text{ for some } n_0 \leq m \leq n-1) \\
&\leq |\mathbf{P}(Y(n) = j) - \pi_j| + 2\mathbf{P}(X_2(m) = 0 \text{ for some } n_0 \leq m < \infty)
\end{aligned}$$

We now make use of property (1) to choose n_0 . Since $\lim_{n \rightarrow \infty} X_2(n) = \infty$ almost surely, we have for every j a random time (not a topping time) $\nu(j)$ such that

$$\nu(j) = \sup\{n : X_2(n) < j\},$$

and $\mathbf{P}(\nu(j) < \infty) = 1$. We can now choose n_0 large enough so that

$$\mathbf{P}(\nu(1) \geq n_0) < \epsilon/2$$

Hence, with probability exceeding $1 - \epsilon/2$, $X_2(n) > 0$ for all $n \geq n_0$, and we have

$$|\mathbf{P}(X_1(n) = j) - \pi_j| < |\mathbf{P}(Y(n) = j) - \pi_j| + \epsilon$$

We therefore have:

$$\begin{aligned}
&\limsup_{n \rightarrow \infty} \left(\sup_j |\mathbf{P}(X_1(n) = j) - \pi_j| \right) \\
&< \limsup_{n \rightarrow \infty} \left(\sup_j |\mathbf{P}(Y(n) = j) - \pi_j| \right) + \epsilon
\end{aligned}$$

but

$$\limsup_{n \rightarrow \infty} \left(\sup_j |\mathbf{P}(Y(n) = j) - \pi_j| \right) = \lim_{n \rightarrow \infty} \left(\sup_j |\mathbf{P}(Y(n) = j) - \pi_j| \right) = 0$$

and we have shown that

$$\limsup_{n \rightarrow \infty} \left(\sup_j |\mathbf{P}(X_1(n) = j) - \pi_j| \right) < \epsilon$$

for an arbitrary $\epsilon > 0$. This completes the proof \square

Remark 1. The same scheme works in a more general setting, of a general measurable state space Markov process $\mathcal{X}(t)$, in discrete time, where assumption (1) is replaced by the assumption that some test function $L(\mathcal{X}(t)) \rightarrow \infty$ almost surely as $t \rightarrow \infty$, and assumption (2) says that conditional on $L(\mathcal{X}(t)) > 0$, a process which is a function of $\mathcal{X}(t)$, say $M(\mathcal{X}(t))$, satisfies that $M(\mathcal{X}(t)) | L(\mathcal{X}(t)) = l$ is ergodic, and independent of the value $l > 0$. The conclusion then is that as $t \rightarrow \infty$, the distribution of $M(\mathcal{X}(t))$ converges to the invariant distribution of $M(\mathcal{X}(t)) | L(\mathcal{X}(t)) = l > 0$. A similar conclusion holds in continuous time.

Remark 2. If one thinks of $X_2(n)$ as being an autonomous Markov chain, then condition 1 of the lemma above means that $X_2(n)$ is transient. A natural question is what happens if we replace this condition with the requirement that $X_2(n)$ is null-recurrent.

The results of [11] hold in the case that some of the nodes in a Jackson network have service rates which are exactly equal to the arrival rates. The question arises whether this could be proven for a general two-component state Markov chain, i.e. whether the lemma above is valid if one replaces condition 1 with the convergence of $X_2(n)$ to infinity in distribution.

The proofs in [11] are based on the monotonicity of the network, and the convergence of the distribution of $X_1(n)$ to the natural limiting one will indeed hold for all specific models exhibiting such monotonicity. A relatively general statement (Lemma 2) and its proof are given below.

However, such a convergence does not hold in general, and here is an example.

Example. Let $X_2(n)$ be a simple random walk on \mathbb{Z}^+ reflected at 0:

$$X_2(n) = \max\{0, X_2(n-1) + \xi(n)\},$$

where $\{\xi(i)\}$ are i.i.d with $\mathbf{P}(\xi(1) = 1) = \mathbf{P}(\xi(1) = -1) = 1/2$ and $X_2(0) = 0$.

Define the first moment this random walk returns to 0 as

$$t(1) = \inf\{n \geq 1 : X_2(n) = 0\}. \quad (7)$$

It is known that $t(1) < \infty$ a.s. but $\mathbf{E}t(1) = \infty$. In fact $\mathbf{P}(t(1) = k) \sim Ck^{-3/2}$ as $k \rightarrow \infty$.

Assume now that

$$X_1(n) = \max\{0, X_1(n-1) + \eta(n)\},$$

where

$$\eta(n) = \begin{cases} -1, & \text{if } X_2(n-1) > 0, \\ \psi(n) & \text{if } X_2(n-1) = 0 \end{cases}$$

for an i.i.d sequence $\psi(n)$ that does not depend on the dynamics of $\{X_2(n)\}$ and is such that $\mathbf{P}(\psi(n) > k) \sim k^{-\alpha}$ as $k \rightarrow \infty$, where $-0 < \alpha < 1/2$. The distributions of $t(i)$ and $\psi(i)$ are regularly varying and, therefore, *long-tailed*. Then (see e.g. Chapter 2 of [9]),

$$\mathbf{P}(\psi(1) - t(1) > x) \sim \mathbf{P}(\psi(1) > x) \quad \text{and} \quad \mathbf{P}(t(1) - \psi(1) > x) \sim \mathbf{P}(t(1) > x), \quad (8)$$

as $x \rightarrow \infty$.

The evolution of $X_1(n)$ is rather simple: the value is decremented by 1 at each step while $X_2(n)$ is positive and jumps up by a random variable with a distribution of $\psi(1)$ when $X_2(n)$ hits zero. It is clear that, conditioned on $X_2(n)$ staying always positive, $X_1(n)$ converges to 0, regardless of its starting point. We are going to show now that, unconditionally, $X_1(n)$ converges to infinity a.s. For that, it is enough to show that random variable

$$\tau = \inf\{n \geq 1 : X_1(n) = 0\}$$

is improper, i.e. $\mathbf{P}(\tau = \infty) > 0$.

Along with (7), define

$$t(k+1) = \inf\{n > t(1) + \dots + t(k) : X_2(n) = 0\} - \sum_{j=1}^k t(j), \quad \text{for } k \geq 1.$$

Define now a random walk

$$S_n = \sum_{i=1}^n (\psi(i) - t(i)).$$

Assume that $X_1(0) = X_2(0) = 0$. Then $\zeta(i) := \psi(i) - t(i)$ are i.i.d. random variables, with $\mathbf{E}\zeta^+ = \mathbf{E}\zeta^- = \infty$ where $\zeta^+ = \max(\zeta(1), 0)$ is the positive part and $\zeta^- = \max(-\zeta(1), 0)$ the negative part of random variable $\zeta(1)$.

It is clear that

$$\mathbf{P}(\tau_1 = \infty) = \mathbf{P}(X_1(n) > 0 \text{ for all } n > 0) = \mathbf{P}(S_n > 0 \text{ for all } n > 0)$$

and it is sufficient to show that $S_n \rightarrow \infty$ a.s. For this we apply Theorem 2 of [8]. In order to be consistent with the notation of the paper, we let, for $x > 0$,

$$m_+(x) = \mathbf{E} \min(\zeta^+, x).$$

By Theorem 2 of [8], $S_n \rightarrow \infty$ a.s. (and then $S_n/n \rightarrow \infty$ a.s.) if and only if

$$J_- = \mathbf{E} \frac{\zeta^-}{m_+(\zeta^-)} < \infty.$$

By (8), $m_+(x) \sim \mathbf{E} \min(\psi(1), x) \sim x^{1-\alpha}/(1-\alpha)$ as $x \rightarrow \infty$ and, further,

$$J_- \approx (1-\alpha) \mathbf{E} \frac{t(1)}{t(1)^{1-\alpha}} = (1-\alpha) \mathbf{E} t(1)^\alpha < \infty.$$

Remark 3. It is clear that the result of the lemma above holds if one replaces its condition 2 by the requirement that $\mathbf{P}(X_1(n+1) = j | X_1(n) = i, X_2(n) = l) = P_{i,j}$, for all values of $l = (l_1, \dots, l_m)$ with all $l_i > N$, for any fixed N .

An interesting question is to consider the case when

$$\mathbf{P}(X_1(n+1) = j | X_1(n) = i, X_2(n) = l) \rightarrow P_{i,j}$$

when $l \rightarrow \infty$, i.e. making the dynamics of $X_1(n)$ asymptotically independent of the position of $X_2(n)$ rather than simply independent of it. This question requires further efforts, and we plan to pursue this research direction.

Lemma 2. Assume again that $X(n) = (X_1(n), X_2(n))$ is a Markov chain taking values in \mathbb{Z}_+^{k+m} . Suppose that Condition 2 of Lemma 1 continues to hold, while Condition 1 is replaced by $\tilde{1}$. $X_2(n) \rightarrow \infty$ in probability (again, coordinate-wise), given $X_1(0) = 0$ and $X_2(0) = 0$.

Further, assume that

3. Markov chain $X(n)$ is monotone, in the following sense. For $y \in \mathbb{Z}_+^{k+m}$, let $C_y = \{z \in \mathbb{Z}_+^{k+m} : z \geq y\}$ where \geq is the standard partial ordering in \mathbb{Z}^{k+m} . Then monotonicity means that

$$\mathbf{P}(X(1) \in C_y | X(0) = x) \geq \mathbf{P}(X(1) \in C_y | X(0) = \hat{x}),$$

for all $x \geq \hat{x}$ and y from \mathbb{Z}_+^{k+m} .

Then the conclusion of Lemma 1 holds again.

Proof. Again, we consider the case $k = m = 1$ only. In what follows, all equalities and inequalities hold a.s.

It is known [6] that a Markov chain may be represented as a stochastic recursion

$$X(n+1) = f(X(n), U_n) \equiv (f_1(X(n), U_n), f_2(X(n), U_n)) = (X_1(n+1), X_2(n+1))$$

where U_n are i.i.d. random variables having the uniform-(0,1) distribution. Also, monotonicity if $X(n)$ implies that f, f_1, f_2 may be chosen monotone in their first argument. We assume U_n to be given for all $-\infty < n < \infty$.

By the lemma conditions there exists a stationary Markov chain $\widehat{X}_1(n)$, $-\infty < n < \infty$ with distribution $\pi = \{\pi_j\}$ that is measurable with respect to $\{U_n\}$ and satisfies recursion

$$\widehat{X}_1(n+1) = f_1((\widehat{X}_1(n), 1), U_n), \quad -\infty < n < \infty.$$

Also, for any $l = 0, 1, 2, \dots$, one may introduce a Markov chain $\widehat{X}_1^{(-l)}(n)$, $n = -l, -l+1, \dots$ that starts from $\widehat{X}_1^{(-l)}(-l) = 0$ at time $-l$ and satisfies recursion $\widehat{X}_1^{(-l)}(n+1) = f_1((\widehat{X}_1^{(-l)}(n), 1), U_n)$. By the monotonicity,

$$0 \leq \widehat{X}_1^{(-1)}(0) \leq \widehat{X}_1^{(-2)}(0) \leq \dots \leq \widehat{X}_1(0)$$

and, moreover, there is an a.s. finite time ν (called *backward coupling time*, see [6]) such that

$$\widehat{X}_1^{(-l)}(0) = \widehat{X}_1(0), \quad \text{for all } l \geq \nu.$$

The latter follows from the uniqueness of the stationary distribution and from the fact that monotone convergence on the lattice must be with coupling (see again [6] for supportive arguments).

Further, for any $l = 1, 2, \dots$, consider an auxiliary Markov chain $X^{(-l)}(n) = (X_1^{(-l)}(n), X_2^{(-l)}(n))$, $n = -l, -l+1, \dots$ that starts from $X^{(-l)}(-l) = (0, 0)$ at time $-l$. Then, by the monotonicity, for any time $n \leq 0$ and for any $-l_1 \leq -l_2 \leq n$,

$$X^{(-l_1)}(n) \geq X^{(-l_2)}(n)$$

and also

$$\widehat{X}_1(n) \geq X_1^{(-l_1)}(n) \geq X_1^{(-l_2)}(n).$$

Then

$$X_2^{(-l)}(n) \uparrow \infty \quad \text{a.s. as } l \rightarrow \infty. \quad (9)$$

for any fixed $n \leq 0$. The sample-path monotonicity in (9) follows from the fact that the Markov chains start from the minimal state and from the induction arguments. Indeed, $X^{(-l-1)}(-l) \geq (0, 0) = X^{(-l)}(-l)$ a.s and then, for any $-l \leq n$, if $X^{(-l-1)}(n) \geq X^{(-l)}(n)$ a.s, then

$$X^{(-l-1)}(n+1) = f(X^{(-l-1)}(n), U_n) \geq f(X^{(-l)}(n), U_n) = X^{(-l)}(n+1) \quad \text{a.s.}$$

Then the a.s convergence to infinity in (9) follows from the monotonicity and the convergence to infinity in probability (since $X_2(n)$ and $X_2^{(-l)}(n)$ have the same distribution, for any n and any l).

Take any small $\varepsilon > 0$ and choose $N \gg 1$ such that $\mathbf{P}(\nu > N) \leq \varepsilon/2$ and then $L > N$ such that

$$\mathbf{P}(X_2^{(-L)}(n) \geq 1, \quad \text{for all } -N \leq n \leq -1) \geq 1 - \varepsilon/2.$$

Then, on the event

$$A := \{\nu \leq N\} \cap \bigcap_{n=-N}^{-1} \{X_2^{(-L)}(n) \geq 1\}$$

of probability at least $1 - \varepsilon$, we have that

$$X_1^{(-L)}(n) = \widehat{X}_1^{(-L)}(n), \quad \text{for all } -N \leq n \leq 0$$

and, in particular,

$$X_1^{(-L)}(0) = \widehat{X}_1^{(-L)}(0) = \widehat{X}_1(0).$$

Therefore, by the monotonicity, on the event A we have

$$X_1^{(-l)}(0) = \hat{X}_1(0),$$

for all $l \geq L$.

Now start Markov chain $X(n)$ from $X(0) = (0, 0)$ at time 0. Then, for any $l \geq L$ and for any set B ,

$$|\mathbf{P}(X_1(l) \in B) - \pi(B)| = |\mathbf{P}(X_1^{(-l)}(0) \in B) - \pi(B)| \leq \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we may conclude that the distribution of $X_1(n)$ converges to π in the total variation if the initial value is $(0, 0)$.

If the Markov chains starts now from another initial state, say $X(0) = (i, j)$, then it may be squeezed between $X_1(n)$ that starts from 0 and, say, sequence $\tilde{X}_1(n+1) = f_1(\tilde{X}_1(n), 1), U_n$ that starts from $\tilde{X}_1(0) = i$. Since both boundary sequences couple with the stationary sequence $\hat{X}_1(n)$, the result follows. □

Remark 4. In Condition 3 of Lemma 2, we require monotonicity in *both* components, X_1 and X_2 . Because of that, convergence to infinity in (9) is monotone too (and, therefore, it occurs almost surely). Condition 3 is satisfied for Jackson networks.

However, this condition was taken just to make the proof simpler. One may weaken the condition, by assuming monotonicity in the first component only. Then convergence in (9) holds in probability, but the statement of the lemma continues to hold – one needs a slightly more detailed analysis.

References

- [1] I.J.B.F. Adan, G. Weiss (2011) Exact FCFS matching rates for two infinite multi-type sequences. *Operations Research* 60:475-489, 2012
- [2] I.J.B.F. Adan, G. Weiss (2014) A queue with skill based service under FCFS-ALIS: steady state, overloaded system, and behavior under abandonments, *Stochastic Systems*, 4(1):250-299.
- [3] I.J.B.F. Adan, A. Busic, J. Mairesse, G. Weiss (2015) Reversibility and further properties of FCFS infinite bipartite matching, arXiv:1507.05939v1 [math.PR].
- [4] A. Aziz, V. Shneer, P. Thiran (2013). Wireless Multi-hop Networks Beyond Capacity. *The 19th IEEE International Workshop on Local and Metropolisian Area Networks*.
- [5] R. Boorstyn, A. Kershbaum, B. Maglaris, V. Sahin (1987). Throughput analysis in multi-hop CSMA packet radio networks. *IEEE Transactions on Communications*. Vol. 35, 267–275.
- [6] A.A. Borovkov, S.G. Foss (1992). Stochastically recursive sequences and their generalisations. *Siberian Advances in Mathematics*, Vol. 2, 16–81.
- [7] D. Denteneer, S. Borst, P. van de Ven, G. Hiertz (2008). IEEE 802.11s and the Philosophers’ problem. *Statistic Neerlandica*. Vol. 62, 3, 283–298.
- [8] K.B. Erickson (1973) The strong law of large numbers when the mean is undefined Trans. American Mathematical Society 185:371–381.

- [9] S.Foss, D. Korshunov and S. Zachary. An Introduction of Heavy-Tailed and Subexponential Distributions. 2nd Edition, Springer, 2013.
- [10] S. Foss, N. Chernova and A. Kovalevskii (1996) Stability of Polling Systems with State-Independent Routing *Proceedings of 34th Allerton Conference*, Monticello, Illinois, 1996, 220–227.
- [11] J.B. Goodman and W.A. Massey (1984) The Non-Ergodic Jackson Network, *J Applied Probability* 21(4):860–869.
- [12] V. Shneer, P.M. van de Ven (2011). Comparing slotted and continuous CSMA: throughputs and fairness. *SIGMETRICS Performance Evaluation Review*. **39**, 2, 35–37.
- [13] V. Shneer, P.M. van de Ven (2015). Stability and instability of individual nodes in multi-hop wireless CSMA/CA networks. *SIGMETRICS Performance Evaluation Review*. **43**, 2, 19–21.